Predicting Cross-Country Skiing FIS Points with Taper Load Sequences and Neural Networks

Joonas Pääkkönen*

*Department of Informatics, Dalarna University, Borlänge, Sweden + email address: jpa@du.se

Abstract

We investigate the training diary dataset of a former elite female cross-country skier who won, e.g., a gold medal in the 2014 Olympic relay. The dataset spans from spring 2012 through spring 2018 with a yearly average of 708.95 training hours. These 6 years cover 119 competitions averaging 69.29 FIS points. We use a multilayer perceptron (MLP) with rectified linear units and 28-day taper load sequence (28-TLS) inputs to predict FIS points. Plots demonstrate that averaged MLP outputs yield promisingly accurate predictions. We also present taper load sequences that correspond to select realised and MLP-predicted FIS point minima. The timing error between the predicted minimum and the realised interval start minimum is only 3 days over a time span of more than 4.5 years. According to our results, the problem of FIS point prediction appears complex, yet tractable through a simple neural network.

1 Introduction

The planning of professional-level cross-country skiing training is heavily results-oriented. It is therefore of extreme importance to understand the relationship between training stimuli and competition performance outcomes. Training program planning largely boils down to understanding the timing of training session modes, loads, intensities and frequencies such that the expected performance level is maximised at competition time. This especially applies to the sensitive time periods close to important competitions, commonly referred to as taper periods. Such periods typically involve significant training load reductions.

One approach to training program planning is through mathematical fitness-fatigue models (FFMs) discussed in, e.g., [1, 2, 3, 4] and the references therein. However, such models have traditionally relied on explicit, straightforward formulae, elementary differential equations or state-space representations, which may lead to overly simplified models. One might argue that artificial neural networks offer a more intriguing approach due to their ability to capture more complicated, and non-linear, relationships between inputs and outputs. We argue that there is a non-linear relationship between competition results and training loads, and that this is due to the fact that competition results do not indefinitely improve or deteriorate with training load – once the training load exceeds or drops below certain tipping point threshold values, competition performance is expected to drastically decline. One might hence argue that for any given competition, there exists a non-trivial, non-zero sequence of daily taper loads that yields the best expected performance outcome.

In this paper, we use an artificial neural network model with rectified linear unit activation functions to predict cross-country skiing competition performance with taper load sequences. Here a taper load sequence is defined as an ordered tuple of daily training time-durations (loads, in hours) of consecutive days. Our work largely builds around the ideas discussed in [5], where a neural network was successfully applied to
a swimming result prediction problem, which inspires further efforts in building performance prediction models for training program planning. While the training characteristics of highly successful cross-country skiers have been extensively reported [6, 7, 8], to the best of our knowledge this is the first work that applies artificial neural networks to predict cross-country skiing performance. In this work, performance is measured by the widely-used Fédération Internationale de Ski (FIS) points [9].

2 Training Diary Dataset

We study the training diary of a former elite female cross-country skier. The whole training diary spans over 6 years (2,191 days) from April 23, 2012 to April 22, 2018 (ages 26–32) and covers over 2,600 training sessions and over 4,200 training hours with an average weekly training duration (load) of 13.60 hours (= 7 × the daily mean duration of 1.94 hours) of which 90.33% is low intensity training (LIT, training zones A1–A2, 60–84% of maximum heart rate), while the rest of it, 9.67%, is high intensity training (HIT, training zones A3–A3+, ≥ 85% of maximum heart rate). Over these 6 years, the skier attended 119 FIS point competitions (68 interval starts, 23 mass starts, 17 pursuits, 11 skiathlons) that produced an average of 69.29 points with a minimum of 6.90, maximum of 207.95 points and point quartiles $Q_1 = 45.36$, $Q_2 = 61.51$, and $Q_3 = 84.21$.

Figure 1 presents all 119 FIS points (blue circles), 28-day trailing average weekly loads (28-TAWL, red line), over all 6 competition season spans (grey). For example, on day 1680, the skier was carrying a training load of approximately 15 hours of weekly training from the past 4 weeks (28 days). Note that Fig. 1 presents no predictions, only realisations, and that there are essentially two figures in one figure: the scatterplot of all the FIS points and the line plot of TAWLs. A competition season span is here defined as the day index span from the first FIS competition of a competition season to the last one.

Figure 1: All 119 FIS point realisations (blue circles), 28-TAWLs (red line), and competition seasons (grey).
Fig. 1 shows that for each of the 6 training seasons, the highest 28-TAWL was over 20 hours, peaking at over 23 hours, while the 28-TAWLs dropped to as low as 7–10 hours in the competition seasons with an average daily load of 1.70 hours over all 28-day taper period days. Over all the taper periods, 78.89% of all training was specific training (on-snow skiing and roller skiing, i.e., not running, cycling, or ski-walking). Of the specific training in taper periods, 96.30% was on-snow skiing and 3.70% was roller skiing, while 78.86% of non-specific training was running. The LIT/(LIT+HIT) load ratio was 89.49% over all taper period days.

One interesting observation we make is that, e.g., in competition season 6, the lowest FIS points were produced when there was a dip in the 28-TAWL. However, at the end of competition season 6 there was an even deeper dip, yet the lowest FIS points of season 6 were not produced during this deepest dip. This observation supports our assumption that FIS points are not simply directly proportional to TAWLs.

In this paper, we focus on daily loads, i.e., the total daily time durations for which the athlete trained during a day. Moreover, we investigate load sequences, i.e., tuples of loads. We define \( t \)-day trailing average weekly load (\( t \)-TAWL) of any given day as the trailing moving average load of the previous \( t \) days, and \( t \)-day taper load sequence (\( t \)-TLS) of any given day as the ordered sequence (tuple) of loads of the previous \( t \) days.

### 3 Prediction Models

FIS point value \( P \) scored by a skier in a competition is directly proportional to the skier’s time-loss \( T_+ \geq 0 \) to the winner as \( P(T) := (T/T_1 - 1)F = (F/T_1)T_+ \), where \( T = T_1 + T_+ \) is a competitor’s time-result, and \( T_1 \) is the winner’s time-result, and \( F \) is a constant that depends on competition type [9]. Each of the models that we build can predict FIS points for any day \( i > t \) whether or not in reality there was a competition on day \( i \) for any taper period length \( t \) (in days). If the \( m \)-day trailing moving average for the final prediction is desired, we may predict for any index \( i > (t + m) \).

In terms of random variables, we study the relative effect of load predictor \( X \) on FIS point response variable \( P \), and thus build regression models to predict \( Y = \log P \), where \( \log(\cdot) \) is the natural logarithm. The log-transformation dampens the effect of large absolute training errors on large target variable values, including outliers. Also, we are especially interested in models that can predict low FIS point values, and we do not, therefore, punish the models for prediction errors on exceptionally large values of \( P \).

We train two models: the naive LS model and the more complicated MLP model, both with 28-day taper period load data, so that the mean squared logarithmic error (MSLE) between training set target values (FIS points) and model predictions is minimised. The training set consists of all the 24 FIS point observations of season 1 competitions (13 interval starts, 5 mass starts, 3 pursuits, 3 skiathlons).

### 3.1 LS Model

We define the naive load scaler (LS) model as \( \hat{y}_i(x_i) := cx_i \), where scalar \( c = (x^T x)^{-1}x^T y \) is derived from the ordinary least squares estimate (OLS) with training set \( t \)-TAWL vector \( x \) and training set vector \( y \) of the corresponding FIS points, \( t \)-TAWL scalar \( x_i = \frac{1}{t} \sum_{j=1}^{t} d_{i-j} \), where scalar \( d_i \) is the total load of day \( i \), and scalar \( \hat{y}_i \) is the logarithm of the FIS point prediction at average load \( x_i \) over \( t \) days prior to a competition on day \( i \). In other words, we fit a logarithmic model \( \log p = y = cx \) of FIS point values \( p \).
3.2 MLP Model

We use MATLAB function fitrnet [10] to fit a (28, 14, 2, 1) fully connected multilayer perceptron (MLP) regression artificial neural network (ANN) with rectified linear unit (ReLU) activation functions \( f(x) = \max(0, x) \) with the Standardize flag set to true. The quadruple (28, 14, 2, 1) refers to 28-day taper load sequence (28-TLS) inputs compressed into 14 neural network inputs, 1 hidden layer with 2 neurons and 1 regression output. We select the ReLU activation function due to the non-linearity between skier training stimuli and competition responses as discussed earlier and also in, e.g., [5], where a hyperbolic tangent activation was used. We argue that ReLU activation functions fit the problem due to their sharp transitions which correspond to the arguably steep degradation of competition performance when the training load greatly deviates from its optimum, i.e., the load or loads that would yield the best competition performance.

MLP inputs before dimensionality reduction consist of \( t = 28 \) daily loads prior to a prediction and the final predictions are trailing moving averages of \( m = 28 \) predictions. Load sequence dimension is reduced from \( t = 28 \) to \( l = 14 \) by selecting the principal component analysis (PCA) scores that correspond to the 14 highest eigenvalues of the sample covariance matrix of the 28-TLSs of all the 2,163 predictable days. These 14 PCA components explain approximately 69.26% of the total load sequence variance.

4 Results

Fig. 2a presents the load scaler (LS) predictions (black line) when season 1 is used for training and seasons 2–6 are used for testing, training set points (blue crosses), test set points (red circles), and competition season spans (grey). The black line of Fig. 2a corresponds to the red line in Fig. 1. Fig. 2b presents the final 28-day trailing moving averages of the multilayer perceptron (MLP) predictions (black line) when season 1 is used for training and seasons 2–6 are used for testing, training set points (blue crosses), test set points (red circles), and competition season spans (grey). The MLP predictions are visibly better than those of the LS model. Table 1 describes sequences prior to realised competitions and the sequence that gives the global MLP-predicted minimum when the search space consists of all possible 56-day sequences of the dataset.

4.1 MLP Predictions

We make several interesting observations regarding the MLP predictions in Fig. 2b where arrows indicate select critical points, such as the MLP-predicted global FIS point minimum on day 2076 and the realised global interval start FIS point minimum on day 2079. While the scaler model of Fig. 2a incorrectly predicts that the best results of season 6 will be produced at the end of the season, the MLP model correctly predicts that the it will occur before the halfway mark of the competition season. The MLP model seems to understand that the 28-TAWL is too low at the end of season 6 to yield low FIS points.

Four lowest FIS points were produced either in a pursuit or a mass start competition and all these four competitions gave clearly better results than the fifth best competition. The fifth best competition was the best interval start competition (10 km classic) on day 2079. This is only 3 days away from the predicted best day on day 2076 despite the prediction being made 1,741 days (> 4.5 years) after the last training set observation of season 1. Note that most of the competitions were interval starts.
Cross-Country Skiing FIS Points

Figure 2: FIS point predictions.

Table 1: TAWLs of days pointed out in Fig. 2b (q.v. Fig. 3).

<table>
<thead>
<tr>
<th>Index $i$</th>
<th>Season</th>
<th>Description</th>
<th>56-TAWL at $i$</th>
<th>28-TAWL at $i$</th>
<th>FIS points</th>
</tr>
</thead>
<tbody>
<tr>
<td>1777</td>
<td>5</td>
<td>realised, global</td>
<td>12.36 hours</td>
<td>12.70 hours</td>
<td>6.90</td>
</tr>
<tr>
<td>2076</td>
<td>6</td>
<td>MLP-predicted, global</td>
<td>12.59 hours</td>
<td>8.64 hours</td>
<td>31.85</td>
</tr>
<tr>
<td>2079</td>
<td>6</td>
<td>realised, interval start</td>
<td>12.46 hours</td>
<td>7.91 hours</td>
<td>25.98</td>
</tr>
<tr>
<td>2083</td>
<td>6</td>
<td>realised, season-specific</td>
<td>11.79 hours</td>
<td>8.85 hours</td>
<td>7.52</td>
</tr>
</tbody>
</table>
4.2 Best Taper Load Sequences

Fig. 3 shows the 56-TLSs that correspond to the TAWLs in Table 1. All these sequences produced or predict low FIS points, and all the corresponding 56-TAWLs are of similar magnitude, while most of the 28-TAWLs are significantly lower than the 56-TAWLs. Further, TAWL drops from approximately 16 hours to 8 hours across the two consecutive 28-day periods as the mean over the 56-TAWLS is approximately 12 hours. On day 2076 there was no competition in actuality, but if there had been, it would have yielded the lowest FIS point realisation according to the MLP model.

Figure 3: 56-TLSs for competitions described in Table 1 and pointed out in Fig. 2b.

5 Conclusions

We have studied the training diary data of a former elite female cross-country skier, and built a multilayer perceptron model with taper load sequence inputs that appears to produce accurate predictions several years into the future. We have presented taper load sequences that correspond to low FIS points, and noted that both the predicted and the realised 28-day trailing average weekly loads typically halve from approximately 16 hours to 8 hours over two consecutive 28-day periods prior to low FIS point competitions. Our findings strongly underline the importance of taper period load sequence planning 28–56 days prior to competitions.
Acknowledgements

The author is deeply indebted to Anna Jönsson Haag for all her input, including her enthusiastic support. The author would also like to thank Martina Höök, Øyvind Karlsson, and Mikael Svarén for inspiring discussions and helpful assistance. Furthermore, the author is affiliated with the Swedish Unit for Metrology in Sports, Department of Medicine, Sport and Fitness Sciences, Dalarna University, Falun, Sweden.

References


